Intron Loss and Gain During Evolution of the Catalase Gene Family in Angiosperms

Julia A. Frugoli,* Mark A. McPeek,* Terry L. Thomas[†] and C. Robertson McClung*

**Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755 and
[†]Department of Biology, Texas A&M University, College Station, Texas 77843*

ABSTRACT

Angiosperms (flowering plants), including both monocots and dicots, contain small catalase gene families. In the dicot, *Arabidopsis thaliana*, two catalase (*CAT*) genes, *CAT1* and *CAT3*, are tightly linked on chromosome 1 and a third, *CAT2*, which is more similar to *CAT1* than to *CAT3*, is unlinked on chromosome 4. Comparison of positions and numbers of introns among 13 angiosperm catalase genomic sequences indicates that intron positions are conserved, and suggests that an ancestral catalase gene common to monocots and dicots contained seven introns. Arabidopsis *CAT2* has seven introns; both *CAT1* and *CAT3* have six introns in positions conserved with *CAT2*, but each has lost a different intron. We suggest the following sequence of events during the evolution of the Arabidopsis catalase gene family. An initial duplication of an ancestral catalase gene gave rise to *CAT3* and *CAT1*. *CAT1* then served as the template for a second duplication, yielding *CAT2*. Intron losses from *CAT1* and *CAT3* followed these duplications. One subclade of monocot catalases has lost all but the 5′-most and 3′-most introns, which is consistent with a mechanism of intron loss by replacement of an ancestral intron-containing gene with a reverse-transcribed DNA copy of a fully spliced mRNA. Following this event of concerted intron loss, the *Oryza sativa* (rice, a monocot) *CAT1* lineage acquired an intron in a novel position, consistent with a mechanism of intron gain at proto-splice sites.

CATALASE (H_2O_2:H_2O_2 oxidoreductase; EC 1.11.1.6) dismutates H_2O_2 into water and oxygen. Together with superoxide dismutase and hydroperoxidase, catalase is part of a defense system for scavenging superoxide radicals and hydroperoxides (Beyer and Fridovich 1987). Catalase is used as a marker for peroxisomes, which are present in almost all eukaryotes (Subramani 1993), and seems to be ubiquitous: no multicellular organism that lacks catalase activity has been found (Scandalios 1987). The active catalase enzyme is a tetrameric iron porphyrin protein. The monomeric catalase subunits are encoded by single genes in most eukaryotes, including mammals (rats, mice, guinea pigs, and humans; Nakashima *et al.* 1989; Quan *et al.* 1986; Shaffer *et al.* 1990; Yaun *et al.* 1996), Drosophila (Orr *et al.* 1990, 1996), and several fungi (GenBank accession number Y07763; Didion and Roggenkamp 1992; Fowler *et al.* 1993; Nakagawa *et al.* 1995). There are, however, two catalase sequences in the nematode *Caenorhabditis elegans* (Waterston *et al.* 1992; Wilson *et al.* 1994), and in the fungi *Saccharomyces cerevisiae* (Cohen *et al.* 1988; Spevak *et al.* 1986), *Candida tropicalis* (Murray and Rachubinski 1987, 1989; Okada *et al.* 1987), and *Aspergillus nidulans* (GenBank accession number U80672; Navarro *et al.* 1996).

In contrast to the situation in most animals and fungi, many plants encode catalase as multigene families (Frugoli *et al.* 1996; Guan and Scandalios 1996; Willekens *et al.* 1994b), which may reflect the multiple and diverse roles played by plant catalases. Plants use catalase in several pathways in addition to those common to other higher eukaryotes. Many oilseed plants store seed energy as lipids, which, upon germination, are converted to sugars through β-oxidation and the glyoxylate cycle (Beevers 1982). The first step in β-oxidation oxidizes a flavin that is subsequently reduced with the concomitant generation of H_2O_2, which then must be broken down by catalase (Trelease 1984). The requirements for rapid and massive flux through β-oxidation and the glyoxylate cycle during germination of oilseed plants suggests that catalase activity should be essential. Catalase is also required in photorespiration, the light-dependent evolution of CO_2 resulting from the oxygenation, as opposed to the carboxylation, of ribulose-1,5-bisphosphate catalyzed by the bifunctional enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Canvin 1990). Catalase is required to dismutate H_2O_2 produced during the peroxisomal oxidation of glycolate to glyoxylate, an intermediate step in the photorespiratory pathway. Photorespiratory catalase activity is essential, and mutants that lack catalase activity are inviable in conditions under which photorespiration occurs (Kendall *et al.* 1983; Somerville and Ogren 1982).

In Arabidopsis, there are three catalase (*CAT*) genes

Corresponding author: C. Robertson McClung, Department of Biological Sciences, 6044 Gilman Laboratory, Dartmouth College, Hanover, NH 03755. E-mail: mcclung@dartmouth.edu

that encode subunits of six to seven detectable tetrameric isozymes (Frugoli *et al.* 1996). We have taken advantage of the power of Arabidopsis as a model system (Meyerowitz and Somerville 1994) to explore the regulation and function of the three catalase genes (McClung 1997). For example, the three *CAT* genes show distinct organ-specific patterns of expression; *CAT1* and *CAT3* mRNAs are most abundant in bolts and leaves, whereas *CAT2* mRNA is most abundant in leaves. All three subunit mRNAs, as well as multiple catalase isozymes, however, are detected in each organ examined, from two isozymes in roots to as many as seven in flowers (Frugoli *et al.* 1996). The three Arabidopsis *CAT* genes also respond differently to light: when dark-adapted plants are returned to light, *CAT1* mRNA is weakly induced and *CAT2* mRNA is strongly induced, whereas *CAT3* mRNA is not induced (E. L. Connolly, H. H. Zhong, R. M. Learned and C. R. McClung, unpublished observations). Furthermore, the expression of two of the *CAT* genes is gated by the circadian clock to distinct times of day: mRNA abundance is maximal at dawn for *CAT2* and is maximal at dusk for *CAT3* (Zhong and McClung 1996; Zhong *et al.* 1994). The expression of the rhythm in *CAT3* mRNA abundance in extended dark is regulated by signaling through both phytochrome and cryptochrome pathways (Zhong *et al.* 1997), while *CAT1* mRNA has no apparent circadian rhythm in abundance (Frugoli *et al.* 1996).

In this study, we address the evolutionary relationships among these three Arabidopsis catalase genes. Phylogenetic analysis based on the amino acid sequence of catalase has suggested that two major groups of catalases are derived from different prokaryotic ancestors, and that plant catalases arose independently of animal and fungal catalases (von Ossowski *et al.* 1993). We focus on angiosperm catalases and provide a phylogenetic analysis that considers more than twice as many catalases as previous analyses have (Guan and Scandalios 1996; Willekens *et al.* 1994a). We suggest that the evolution of this multigene family entailed a series of gene duplications of an ancestral angiosperm catalase that, before the divergence of monocots and dicots, contained seven introns. In Arabidopsis, analysis of sequence similarity and genetic linkage allows us to postulate a sequential order of duplication events during the evolution of the *CAT* gene family, as well as a pattern of intron loss consistent with this order of events. The pattern of intron loss seen in the monocot lineage suggests a mechanism of concerted intron loss that is concordant with the replacement of an ancestral intron-containing gene with a reverse-transcribed DNA copy of a fully spliced mRNA, consistent with the mechanism suggested by Baltimore (1985) and Fink (1987). After this event of concerted intron loss, one rice catalase acquired an intron in a novel position. Comparison of sequences surrounding the site of intron addition within this clade

supports a mechanism, first proposed by Dibb and Newman (1989), of intron gain at proto-splice sites.

MATERIALS AND METHODS

Plant growth conditions: *Arabidopsis thaliana* and *A. griffithiana* plants were grown in constant light (130 μmol·m^{-2}·s^{-1} photosynthetically active radiation) at 20°, harvested, frozen in liquid nitrogen, and stored at −80°.

Southern analysis: Southern analysis was by standard protocols (Ausubel *et al.* 1997) using Nytran Plus (Schleicher & Schuell, Keene, NH) or Hybond-N+ (Amersham, Arlington Heights, IL) membranes. Hybridization and washes were as described (Choi *et al.* 1995). Probes were made by excision of gene-specific DNA fragments (Frugoli *et al.* 1996) from agarose gels after electrophoresis, purification by Qiaquick gel purification kits (Qiagen, Chatsworth, CA), and random primer labeling with [α-^{32}P] dATP using the Klenow fragment of DNA polymerase I (Feinberg and Vogelstein 1984). Membranes were then wrapped in Saran wrap, and autoradiographs were generated on Bio-Max film (Eastman Kodak, Rochester, NY) in autoradiography cassettes at −80° for 1–3 days, depending on the blot.

Bacterial artificial chromosome (BAC) analysis: Two sets of Texas A&M University BAC filters (Choi *et al.* 1995) were hybridized with probes for *CAT1* and *CAT3* (Frugoli *et al.* 1996), and four clones that hybridized to both genes (T10F14, T5B3, T2C19, and T4J5) were analyzed further. The BAC DNA was isolated and digested with *Not*I to release the inserts. CHEF gel analysis (Choi *et al.* 1995) indicated that the inserts ranged from ∼40 kb to >150 kb in size. The 40-kb insert of clone T10F14 was chosen for further analysis.

CAT1 genomic sequence: Restriction fragments from TAMU BAC T10F14 digested either with *Bam*HI and/or with *Xba*I were subcloned (Ausubel *et al.* 1997) into pBluescript KS (Stratagene, La Jolla, CA). The *CAT1*- and *CAT3*-hybridizing fragments were sequenced by dideoxy chain termination using the ABI Dye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA) and various primers designed within the gene and the vector polylinker. The reaction products were run on ABI model 373A and 377 sequenators, and sequences were viewed with ABI Prism View software (Applied Biosystems). The sequenced fragments of clones from both digests were assembled with Geneworks (IntelliGenetics, Mountain View, CA) and various programs of the GCG package (version 8; Genetics Computer Group, Madison, WI), and the resulting sequence was deposited in GenBank under accession number AF021937.

Polymerase Chain Reactions: For confirmation of the *CAT3*/*CAT1* locus structure, 100 ng of DNA from each BAC was used in a 50-μl PCR reaction with 25 nm each of primer 1/3 (5′-ATGGATCCATGATGCTTGAAGAC-3′, corresponding to nt 69–91, using the numbering scheme of GenBank accession number U43340, or nt 4349–4327 according to the numbering scheme of GenBank accession number AF021937), and 3/1 (5′-AAGGATCCTCACATGTGTTGTGT-3′, corresponding to nt 3806–3828, using the numbering scheme of GenBank accession number AF021937), 100 μM dNTPs, 2.5 units of *Taq* polymerase, and 3 mm MgCl$_2$ in PCR reaction buffer (Promega, Madison, WI). Reaction steps were 5 min initial denaturation at 94°, followed by 35 cycles, each of 1 min at 94°, 30 sec at 55°, and 1 min at 72°, and a final elongation step of 5 min at 72°.

We wished to determine whether the absence of the last intron in *CAT1* was conserved among ecotypes of *A. thaliana* (collected from North America, Europe, and Africa), as well as *A. griffithiana*, a related species from Tajikistan (Asia). The

TABLE 1

Primers used in determining *H. vulgare* genomic sequences

Primer name	Primer sequence	Position
BCAT1-1F	5′ GCCATGGATCCCTACAAG 3′	nt 74–91[a]
BCAT1-1R2	5′ TCTGAACCCCAGGAGCCCGGA 3′	nt 365–345
BCAT1-2F	5′ GTTCAGACCCCGGTTATTGTC 3′	nt 359–379
BCAT1-2R2	5′ TTGTGGTGAGCACATTTCGGG 3′	nt 1197–1177
BCAT1-3F	5′ CTCACCACAACAACCATCATG 3′	nt 1188–1208
BCAT1-3R2	5′ TCACCAGCCTGCTTGAAATTG 3′	nt 1374–1354
BCAT1-4F	5′ GCTGGTGAGAGATATCGGTCC 3′	nt 1367–1387
BCAT1-4R2	5′ CCCTCCTCCTTGTCCATCTTA 3′	nt 1573–1553
BCAT2-1F	5′ CTACTTCACTCACTCGAGG 3′	nt 36–54[b]
BCAT2-1R2	5′ GATCACGGTGGAGAAGCGGAC 3′	nt 379–359
BCAT2-2F	5′ GTCCGCTTCTCCACCGTGATC 3′	nt 359–379
BCAT2-2R2	5′ CGTAGTGGTTGTTCTTGAAGC 3′	nt 1190–1169
BCAT2-3F	5′ CTACGACGGCGCCATGAACTT 3′	nt 1186–1206
BCAT2-3R2	5′ GGCTGCACGAAGTCGTTCTCC 3′	nt 1356–1336
BCAT2-4F	5′ AGGTACCGCAGCTGGGCGCCC 3′	nt 1364–1383
BCAT2-4R2	5′ CCATGGCAGATCCATCCATCA 3′	nt 1561–1541

[a] *HvCAT1* primer pairs numbered according to the numbering scheme of GenBank accession number U20777.

[b] *HvCAT2* primer pairs numbered according to the numbering scheme of GenBank accession number U20778.

ecotypes (geographic origin in parentheses) for which data are presented include Be-0 (Germany), Bu-0 (Germany), Col-2 (United States), Est-0 (Russia), La-0 (Poland), Mh-0 (Poland), Ws-0 (Russia), and Ler-0, a laboratory strain derived from La-0 that contains a mutation at the *erecta* locus. In addition, we obtained similar results for the following ecotypes: Bu-0 (Germany), Cvi-0 (Cape Verde Islands), Le-0 (Netherlands), Ms-0 (Russia), Nd-0 (Germany), No-0 (Germany), Po-1 (Germany), RLD-1 and Sei-0 (Italy). These ecotypes originated at altitudes of 1–300 meters (Arabidopsis Biological Resource Center Catalog). For each ecotype or species, 50 ng of genomic DNA was amplified using 25 nm each of primers CAT1-11 [5′-GCGATATC-GGTCAATTACTTCCCTT CAAGG-3′, nt 1248–1269, using the numbering scheme of GenBank accession number U43340 (note that the first 8 nt do not correspond to the Arabidopsis genomic sequence and include an *Eco*RV site added to the primer to facilitate subsequent cloning steps)] and CAT1-12 [5′ GAGATGAATTCATT CAGAAGTTTGGCC-3′, nt 1573–1547, using the numbering scheme of GenBank accession number U43340 (note that mutations A1565T and T1566A have been included in the primer to yield an *Eco*RI site to facilitate subsequent cloning steps)], 100 μm dNTPs, 2.5 units of *Taq* polymerase, and PCR reaction buffer (Invitrogen, San Diego, CA). Reaction steps were 5 min initial denaturation at 92°, followed by 35 cycles, each of 30 sec at 92°, 30 sec at 50°, and 3 min at 72°, and a final elongation step of 5 min at 72°. Reaction products were analyzed by agarose gel electrophoresis.

For determination of *Hordeum vulgare CAT1* and *CAT2* intron positions, a series of primer pairs (Table 1) were designed to amplify, in a set of overlapping products, the complete genomic sequences of the two *CAT* genes from *H. vulgare* cv. Harrington genomic DNA (a gift from J. Sherwood). Amplification reactions included 50 ng barley genomic DNA, 100 μm dNTPs, 25 nM of each primer, 2.5 units of *Taq* polymerase, and PCR reaction buffer (Boehringer-Mannheim, Indianapolis, IN). Reaction steps were 5 min initial denaturation at 94°, followed by 25 cycles, each of 1 min at 94°, 1 min at

55°, and 1 min at 72°, and a final elongation step of 7 min at 72°. PCR products were either subcloned into plasmid pCR2.1 using the TA Cloning Kit (Invitrogen) and then sequenced as described above, or they were directly sequenced, following purification through Centri-Spin 40 columns (Princeton Separations, Adelphia, NJ). We were unable to amplify any *CAT*-related product in reactions using primers BCAT1-1F or BCAT2-1F, so we did not obtain a sequence that spanned the first or second intron for either *H. vulgare CAT* gene. Sequences of *H. vulgare CAT1* and *CAT2* were assembled and analyzed using the GCG programs, and they have been deposited in GenBank under accession numbers AF021938 and AF021939, respectively.

Phylogenetic analysis: We generated a hypothesis for the phylogenetic relationships among the gene sequences from the deduced amino acid sequences of Arabidopsis catalases and of other plant catalases in the GenBank database using cladistic analyses implemented by PAUP version 3.1.1 (Swofford 1993).

RESULTS

Physical characterization of the *CAT3-CAT1* loci: As part of a comprehensive analysis of the catalase gene family of Arabidopsis, we wished to characterize the genomic structure of each of the three *CAT* genes. Low-resolution mapping showed that *CAT1* and *CAT3* are tightly linked on chromosome *1*, whereas *CAT2* is on chromosome *4* (Frugoli *et al.* 1996). To refine the relative chromosomal map positions of *CAT1* and *CAT3* and to determine the genomic structures of these two loci, we identified four BAC clones that hybridized to both *CAT1* and *CAT3*. CHEF gel analysis determined that clone T10F14 contained an insert of only 40 kb (data not shown), so further analysis focused on this
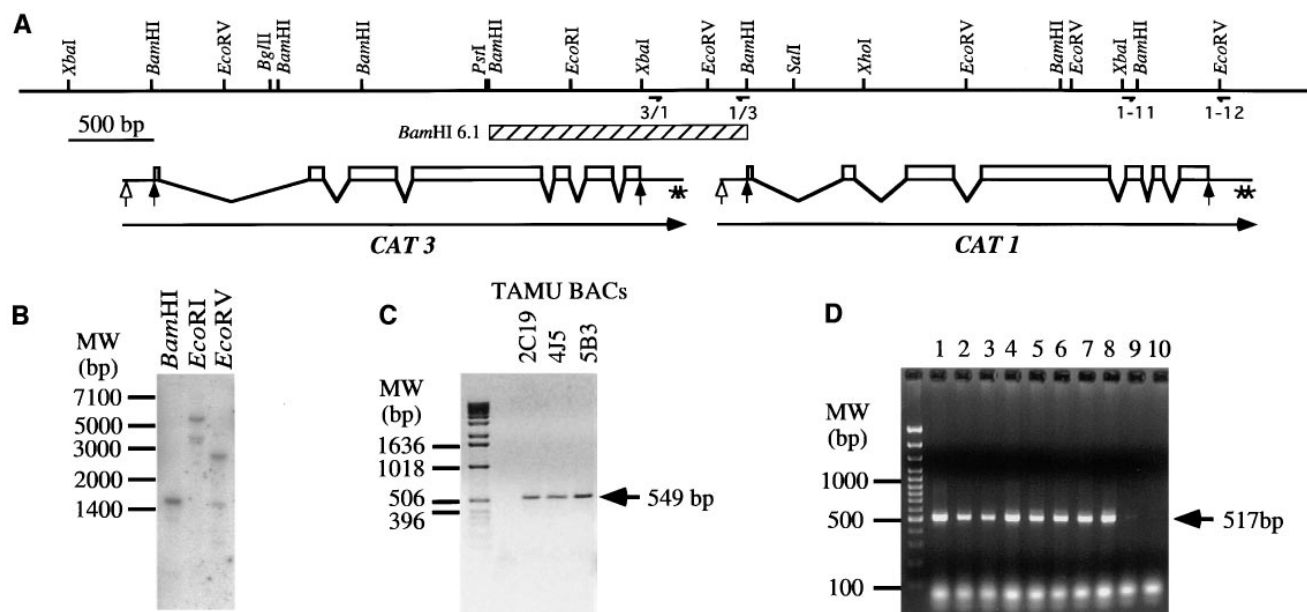
Figure 1.—Physical characterization of the Arabidopsis *CAT3/CAT1* locus. (A) A restriction map of the *CAT3/CAT1* locus is presented in the upper line. The location of primers used in PCR (see text) are indicated by half arrows below the restriction map. The *Bam*HI 6.1 fragment (1555 nt) used as the hybridization probe in Southern analysis is indicated by the hatched box. Intron/exon structure of the two genes is indicated below the restriction map. Open boxes denote coding sequences, horizontal lines indicate 5′ and 3′ untranslated regions, and bent lines indicate introns. Translation start and stop sites are indicated by filled arrows, and transcription starts are indicated by open arrows. Asterisks indicate polyadenylation sites. Both genes are transcribed from left to right, as indicated by the arrows underneath the cartoon. Note that orientation of this locus relative to the chromosome *1* centromere remains uncertain. (B) Southern blot of Columbia ecotype genomic DNA probed with fragment *Bam*HI 6.1 shows that the tight linkage of *CAT3* and *CAT1* seen in TAMU BAC T10F14 is conserved in genomic DNA and therefore does not represent a cloning artifact. (C) Results of PCR amplification of BAC DNA using primers 3/1 and 1/3 (see A) show that the tight linkage of *CAT3* and *CAT1* seen in TAMU BAC T10F14 is conserved in other BACs and therefore does not represent a cloning artifact. (D) Conservation of *CAT1* intron/exon structure across Arabidopsis ecotypes and species. Agarose gel electrophoresis of PCR amplification products using various Arabidopsis ecotypes as templates and primers 1–11 and 1–12 (see A) demonstrates that loss of intron 7 from *CAT1* is conserved among Arabidopsis ecotypes (see materials and methods) and in the related species *A. griffithiana*. Lane 1, *A. griffithiana*; 2, BE-O; 3, BU-O; 4, EST; 5, LA-O; 6, MH; 7, Ws; 8, Ler; 9 and 10, no template DNA.

BAC. Restriction endonuclease digestions and subsequent DNA sequence analysis of BAC T10F14 generated the map of the loci shown in Figure 1A. *CAT1* and *CAT3* are immediately adjacent, with the most distal polyadenylation site of *CAT3* (three sites at nt 3945, 3956, and 3959; GenBank accession number AF021937) only 277 bp upstream from the most proximal site of start of transcription of *CAT1* (five sites at nt 4236, 4238, 4241, 4244, and 4246; GenBank accession number AF021937), which we determined by primer extension (data not shown). The two genes are transcribed in the same direction, although at this time we remain uncertain as to the orientation of the two genes relative to the telomere and centromere of the upper arm of chromosome 1.

We also sought to confirm that the tight linkage of *CAT1* and *CAT3* seen in BAC T10F14 did not reflect some rearrangement that occurred during the cloning procedure. First, we performed Southern analysis of Arabidopsis ecotype Columbia genomic DNA using as a hybridization probe; a *Bam*HI fragment subcloned from TAMU BAC T10F14 (*Bam*HI 6.1, 1555 nt, indi-

cated by a hatched line in Figure 1A) that spanned the *CAT3-CAT1* intergenic region (Figure 1B). *Bam*HI 6.1 detected a single *Bam*HI fragment of 1.55 kb in the genomic DNA. In addition, two *Eco*RI fragments of ∼6 and 4 kb and two *Eco*RV fragments of 2.9 and 1.5 kb were detected, consistent with the map in Figure 1A and with the nucleotide sequence reported in GenBank accession number AF021937. Further confirmation was obtained by designing two oligonucleotide primers (Figure 1A and materials and methods), one in a previously sequenced region of *CAT3* (Zhong and McClung 1996) and one in a previously sequenced region of the *CAT1* cDNA (Frugoli *et al.* 1996). These primers were used in PCR reactions that used the other three TAMU BAC clones that hybridized to both *CAT* genes as templates; in each case, we observed the 549-bp product (Figure 1C) predicted by the map (Figure 1A) and by the nucleotide sequence.

Phylogenetic analysis of angiosperm catalases using amino acid sequences: Including the three Arabidopsis catalases, a total of 37 angiosperm catalase sequences, representing 22 species, as well as two Chlamydomonas

TABLE 2

Catalase sequences used in this study

Species Latin binomial (common name)	Catalase	GenBank accession number	
		cDNA	Genomic
Arabidopsis thaliana (thale cress)	1	U43340	AF021937
Arabidopsis thaliana	2	X64271	X94447
Arabidopsis thaliana	3	nd	U43147, AF021937
Chlamydomonas reinhardtii	1	AF016902	nd
Chlamydomonas reinhardtii	2	Y13220	nd
Cucurbita pepo (pumpkin)	1	D55645	nd
Cucurbita pepo	2	D55646	nd
Cucurbita pepo	3	D55647	nd
Glycine max (soybean)		nd	Z12021
Gossypium hirsutum (cotton)	1	X52135	nd
Gossypium hirsutum	2	X56675	nd
Helianthus annuus (sunflower)		L28740	nd
Hordeum vulgare (barley)	1	U20777	AF021938
Hordeum vulgare	2	U20778	AF021939
Ipomoea batatas (sweet potato)		X05549	nd
Lycopersicon esculentum (tomato)		M93719	nd
Nicotiana glutinosa		AF006067	nd
Nicotiana plumbaginifolia	1	Z36975	nd
Nicotiana plumbaginifolia	2	Z36976	nd
Nicotiana plumbaginifolia	3	Z36977	nd
Nicotiana sylvestrus	1	U07626	nd
Nicotiana tabacum (tobacco)	1	U07627	nd
Nicotiana tabacum	2	U93244	nd
Oryza sativa (rice)	1	D26484	D29966
Oryza sativa	2	X61626	nd
Pisum sativum (pea)		X60169	nd
Ricinus communis (castor bean)	1	nd	D21161
Ricinus communis	2	nd	D21162
Soladenella alpina		Z99633	nd
Secale cereale (rye)		Z54143	nd
Solanum melongena (eggplant)		X71653	nd
Solanum tuberosum (potato)	1	U27082	nd
Solanum tuberosum	2	nd	Z37106
Triticum aestivum (wheat)	1	D86327	nd
Triticum aestivum	2	X94352	nd
Vigna radiata (mung bean)		D13557	nd
Zea mays (maize)	1	X12538	X60135
Zea mays	2	X54819	Z54358
Zea mays	3	X12539	L05934

nd, not deposited in GenBank.

catalase sequences, have been deposited in GenBank (as of November 8, 1997; Table 2). The amino acid sequences of these 39 catalases were aligned with the PILEUP program (version 8 Program Manual, GCG). We used the heuristic search algorithm (branch swapping with 20 replicates of random sequence additions) of PAUP ver. 3.1.1 (Swofford 1993) to generate a hypothesis for the phylogenetic relationships among these sequences. One tree of shortest length was found by this analysis (Figure 2A).

We also performed a bootstrap analysis (100 replicates) using this same algorithm. This bootstrap analysis strongly supports the phylogenetic affinities of a num-

ber of species groups (Figure 2B), but this analysis also indicates that the deep branching patterns identified in the tree of Figure 2A are not robust. This is probably a function of the limited number of genes from different putative catalase lineages that can be included in the current analysis. The topology of the trees in Figure 2, A and B, suggests, however, that the gene duplications giving rise to multiple plant catalases occurred before the divergence of monocots and dicots. The results of our analysis are generally consistent with the previous analyses of smaller sets of plant catalases (Guan and Scandalios 1996; Willekens *et al.* 1994a). Analysis of the increased number of grass (the only group of mono-

A

```
        ┌─ C. reinhardtii 1
        └─ C. reinhardtii 2
        ┌─ H. vulgare 1
        ├─ T. aestivum 1
        ├─ Z. mays 1
        └─ O. sativa 2
        ┌─ Z. mays 2
        └─ T. aestivum 2
        ── H. annuus
        ┌─ N. plumbaginifolia 1
        ├─ N. tabacum 2
        ├─ N. glutinosa
        └─ G. hirsutum 2
        ┌─ A. thaliana 2
        ├─ C. pepo 1
        ├─ V. radiata
        ├─ P. sativum
        └─ G. max
        ┌─ N. plumbaginifolia 3
        ├─ S. alpina
        └─ G. hirsutum 1
        ── R. communis 1
        ── A. thaliana 1
        ┌─ H. vulgare 2
        ├─ S. cereale
        ├─ O. sativa 1
        └─ Z. mays 3
        ── R. communis 2
        ── A. thaliana 3
        ┌─ S. tuberosum 2
        ├─ S. tuberosum 1
        ├─ L. esculentum
        └─ S. melongena
        ┌─ N. sylvestris 1
        ├─ N. tabacum 1
        └─ N. plumbaginifolia 2
        ┌─ C. pepo 2
        ├─ C. pepo 3
        └─ I. batatas
```

B

```
  100   ┌─ C. reinhardtii 1
        └─ C. reinhardtii 2
       100┌─ H. vulgare 1
  100    └─ T. aestivum 1
         ── Z. mays 1
         ── O. sativa 2
  99     ┌─ Z. mays 2
         └─ T. aestivum 2
         ── H. annuus
       73┌─ N. plumbaginifolia 1
  100    └─ N. tabacum 2
         ── N. glutinosa
100      ── G. hirsutum 2
  81     ┌─ A. thaliana 2
         └─ C. pepo 1
  61     ┌─ V. radiata
         └─ P. sativum
         ── G. max
  69     ┌─ N. plumbaginifolia 3
64       └─ S. alpina
         ── G. hirsutum 1
         ── R. communis 1
         ── A. thaliana 1
              100┌─ H. vulgare 2
            59 ├─ S. cereale
  100      ├─ O. sativa 1
            └─ Z. mays 3
         ── R. communis 2
         ── A. thaliana 3
          89┌─ S. tuberosum 2
       100  └─ S. tuberosum 1
        96┌─ L. esculentum
  100  └─ S. melongena
        79┌─ N. sylvestris 1
      99  └─ N. tabacum 1
         ── N. plumbaginifolia 2
  100   ┌─ C. pepo 2
        ├─ C. pepo 3
        └─ I. batatas
  61
```
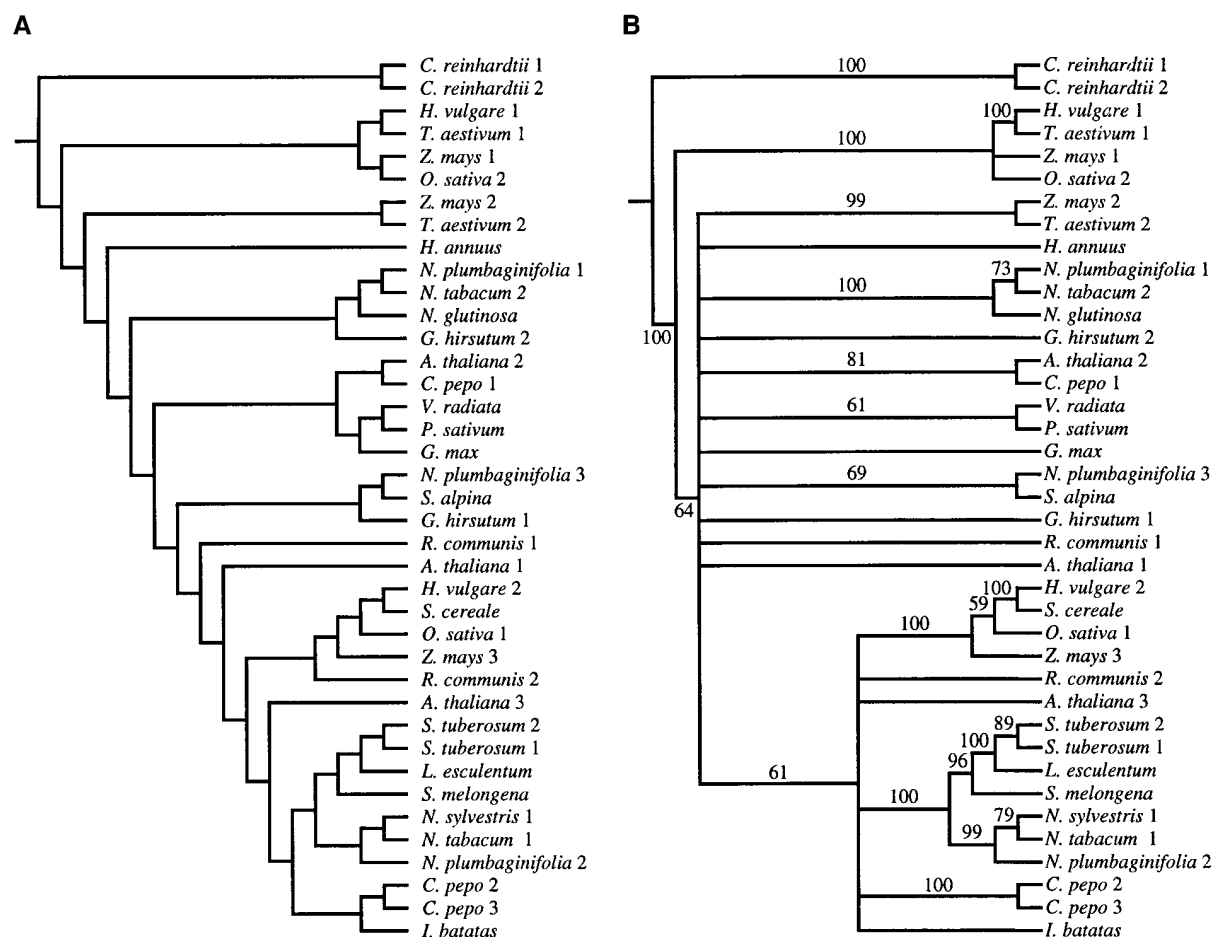
Figure 2.—Phylogenetic relationships among plant catalases in GenBank. (A) The heuristic search algorithm of version 3.1.1 of PAUP (Swofford 1993) using branch swapping with tree bisection-reconnection (the TBR option with 20 replicates of random sequence additions) identified one most parsimonious tree. (B) Bootstrap values resulting from 100 replicates of the heuristic search algorithm (Swofford 1993). GenBank accession numbers for each sequence represented in the trees are given in Table 2.

cots represented) sequences suggests, however, that three subclades of grass catalases exist, each including one of the three known *Zea mays* catalases. Similarly, the three known *Nicotiana plumbaginifolia* catalases define three distinct subclades. One goal of this analysis was to correlate phylogenetic and functional groupings of catalases to provide testable hypotheses concerning the function of the three Arabidopsis catalases and catalases in general. The three Arabidopsis catalases define distinct lineages, but they do not form large subclades with other catalases. Furthermore, the limited number of sequences from representative taxa across the plant kingdom and the paucity of functional analyses of individual plant catalases undermine present attempts to correlate phylogeny and function.

Intron-exon structure of angiosperm *CAT* genes: In addition to the genomic sequences of the three Arabidopsis catalases, genomic sequences were available for eight other catalases (Table 2). For a more complete analysis of the grass lineages, we determined partial genomic sequences for the two barley catalase genes (deposited in GenBank under accession numbers AF021938 and AF021939). Among these 13 sequences, introns were observed at a total of eight positions, which we have numbered according to their position within the genes, with intron 1 closest to the 5' end and intron 8 closest to the 3' end. Numbers, positions, and sizes of the introns are summarized in Table 3 and Figure 3. All the observed introns interrupt the coding sequences, and the positions of the observed introns, relative to the coding sequence, are conserved among all angiosperm catalases, as was noted previously in an analysis of a smaller data set (Guan and Scandalios 1996). Because the majority of *CAT* sequences contain seven introns (1–3 and 5–8) in conserved positions, we suggest that an ancestral catalase gene common to monocots and dicots contained seven introns in these positions, and that the presence of intron 4 at a novel position in the *Oryza sativa CAT1* sequence represents a derived character (see below). Several of the genes lack one or more introns (Table 3 and Figure 3); we suggest that this occurred by intron loss. Arabidopsis *CAT1* lacks intron 8. We confirmed that the absence of this intron in *CAT1* was conserved across ecotypes by PCR analysis

TABLE 3

Positions and lengths of introns in angiosperm catalase genes

Intron number	1	2	3	4	5	6	7	8
Intron position	K5[a]	G38[b]	E130[a]	Q272[a]	E389[a]	K419[a]	R442[c]	Q473[a]
				Intron length (nt)				
Arabidopsis thaliana 1	586	273	178	—	88	91	102	—
Arabidopsis thaliana 2	259	76	78	—	82	98	89	102
Arabidopsis thaliana 3	923	154	91	—	85	100	—	93
Glycine max	91	78	—	—	99	403	98	515
Hordeum vulgare 1	nd	nd	381	—	129	177	103	125
Hordeum vulgare 2	nd	nd	—	—	—	—	—	123
Oryza sativa 1	317	—	—	85	—	—	—	350
Ricinus communis 1	189	235	251	—	237	145	84	100
Ricinus communis 2	733	122	102	—	105	102	—	86
Solanum tuberosum 2	1394	148	91	—	87	87	90	97
Zea mays 1	157	938	679	—	—	163	112	111
Zea mays 2	92	108	—	—	—	109	78	113
Zea mays 3	84	—	—	—	—	—	—	93

[a] Intron follows the codon encoding the indicated amino acid residue, with the numbering-based sequence of *A. thaliana* CAT1, to which all others have been aligned.

[b] Intron inserted between the first and second nucleotides in the codon encoding the indicated amino acid residue.

[c] Intron inserted between the second and third nucleotides in the codon encoding the indicated amino acid residue.

—, intron not present; nd, not determined.

(Figure 1D). Arabidopsis *CAT3* and *Ricinus communis CAT2* lack intron 7. *Glycine max* lacks intron 3. *Z. mays Cat2* lacks both introns 3 and 5. Of particular interest is one subclade of grass catalases (*Z. mays Cat3, O. sativa CAT1*, and *H. vulgare CAT2*), each of which lacks introns

Intron Position

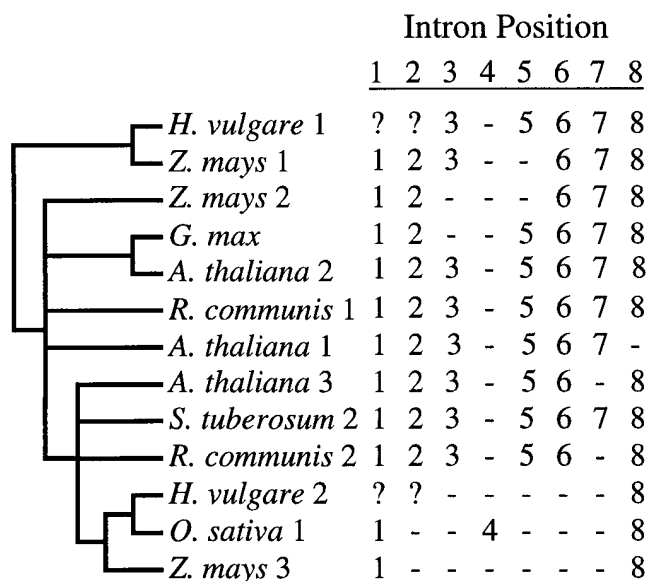|	1	2	3	4	5	6	7	8
H. vulgare 1	?	?	3	-	5	6	7	8
Z. mays 1	1	2	3	-	-	6	7	8
Z. mays 2	1	2	-	-	-	6	7	8
G. max	1	2	-	-	5	6	7	8
A. thaliana 2	1	2	3	-	5	6	7	8
R. communis 1	1	2	3	-	5	6	7	8
A. thaliana 1	1	2	3	-	5	6	7	-
A. thaliana 3	1	2	3	-	5	6	-	8
S. tuberosum 2	1	2	3	-	5	6	7	8
R. communis 2	1	2	3	-	5	6	-	8
H. vulgare 2	?	?	-	-	-	-	-	8
O. sativa 1	1	-	-	4	-	-	-	8
Z. mays 3	1	-	-	-	-	-	-	8

Figure 3.—Most parsimonious tree, generated by pruning the tree of Figure 2A, to include only those genes for which intron information is available. Intron positions are reported in relation to amino acid sequence (see text). Question marks indicate introns whose presence or absence we were unable to determine by PCR.

at positions 3, 5, 6, and 7. The *Z. mays* and *O. sativa* sequences also lack the intron at position 2, although we were unable to amplify the 5′ regions of either of the *H. vulgare* catalases, so we have no information about the presence of introns 1 or 2 for either *H. vulgare* sequence. In addition, the *O. sativa CAT1* sequence contains an intron, intron 4, at a novel position after the codon encoding amino acid residue Q272. These data are summarized in Figure 3, which includes only those sequences for which genomic sequence and, hence, intron positions are known.

DISCUSSION

The majority of studies of plant molecular evolution have focused on the chloroplast genome, and the molecular evolution of plant nuclear genes remains to be comprehensively addressed. Most studies of plant nuclear genes have examined gene families, such as those encoding the small subunit of ribulose-bisphosphate carboxylase/oxygenase (*RBCS*) or alcohol dehydrogenase (*ADH*), in which the protein product of the genes plays a defined biochemical role in a limited set of pathways (Clegg *et al.* 1997). Plant actin genes comprise a large and complex gene family in which diversity in function is paralleled by gene family diversity that exceeds that found in other eukaryotes (Meagher 1991; McDowell *et al.* 1996). Plant catalases play diverse roles in germination, photorespiration, resistance to oxidative stress (McClung 1997), and possibly in mediating

A

tandem duplication on Chr. 1

B

*CAT3**   *CAT1**

duplication and insertion on Chr. 4

C

*CAT3**   *CAT1**   *CAT2*

• loss of intron 7   • loss of intron 8

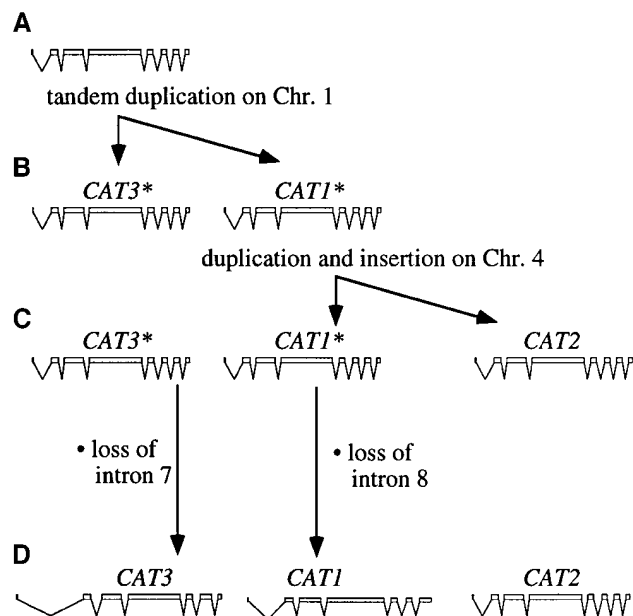**D**

CAT3   *CAT1*   *CAT2*

Figure 4.—A model for the evolution of the catalase three-gene family in Arabidopsis. Asterisks indicate progenitors of present day genes. (A) The ancestral *CAT* gene is postulated to contain seven introns and to reside on chromosome *1*. (B) An initial tandem duplication yielded progenitors of *CAT3* and *CAT1*. (C) A second, more recent duplication of the *CAT1* sequence was associated with the integration of the duplicated copy on chromosome *4*. (D) The loss of intron 7 from *CAT3* occurred after the duplication that originally resulted in the *CAT1* and *CAT3* progenitors, and the loss of intron 8 from *CAT1* followed the second duplication.

signal transduction involving H_2O_2 as a second messenger (Low and Merida 1996; Mehdy *et al.* 1996; Ryals *et al.* 1996; Yang *et al.* 1997). Therefore, the catalase genes, such as the actin genes (Meagher 1991) and the genes of the flavonoid pathway, notably chalcone synthase (Clegg *et al.* 1997), offer a useful system in which to address how the acquisition of multiple metabolic roles influences the evolution of a gene family.

The evolution of multigene families involves multiple mechanisms (Ohta 1991; Fryxell 1996). Any explanation of the evolution of the Arabidopsis *CAT* family must take into account three pieces of data: the greater sequence similarity between *CAT1* and *CAT2* than of either with *CAT3* (Frugoli *et al.* 1996), the tight linkage of *CAT1* and *CAT3* (Figure 1), and the pattern of intron losses (Table 3 and Figure 3). We propose the following sequence of events to account for the evolution of the Arabidopsis *CAT* gene family (Figure 4). The ancestral *CAT* gene, containing seven introns, was one of the two linked genes that now reside on chromosome 1 (Figure 4A), and an initial tandem duplication yielded progenitors of *CAT3* and *CAT1* (Figure 4B). A second, more recent duplication of the *CAT1* sequence was associated with the integration of the duplicated copy on chromosome 4 (Figure 4C). We argue that *CAT1* provided the template for this second duplication event because

CAT1 and *CAT2* are more similar in sequence to each other, at both the nucleotide and amino acid levels, than either is to *CAT3* (Frugoli *et al.* 1996). The loss of intron 7 from *CAT3* occurred sometime after the duplication that originally resulted in *CAT1* and *CAT3*, and the loss of intron 8 from *CAT1* followed the second duplication (Figure 4D).

Interestingly, a second dicot, *Ricinus communis*, also has two tightly linked (by ~2 kb) *CAT* genes that, like Arabidopsis *CAT3* and *CAT1*, are transcribed in the same direction (Suzuki *et al.* 1994). This suggests that the tandem duplication was an early event that occurred before the separation of the Arabidopsis and Ricinus lineages. The upstream genes from each species, Arabidopsis *CAT3* and *R. communis CAT2*, have lost intron 7; we infer that the loss of intron 7 in this gene lineage predated the divergence of the Arabidopsis and *R. communis* lineages. The two downstream genes, Arabidopsis *CAT1* and *R. communis CAT1*, are more related to each other than to the other linked catalase (Arabidopsis *CAT3* and *R. communis CAT2*, respectively). *R. communis CAT1* retains intron 8, however, whereas Arabidopsis *CAT1* has lost intron 8, suggesting that Arabidopsis *CAT1* lost intron 8 after the separation of the lineages. One obvious question is why the tight linkage of these two catalase genes should have persisted over evolutionary time. It is possible that *CAT1* and *CAT3* retain tight linkage to facilitate coregulation, although mRNA expression patterns of these two genes are dissimilar (Frugoli *et al.* 1996). A better test of this hypothesis would be to determine whether CAT1 and CAT3 monomers assemble into mixed tetramers and thus share common functions, which would suggest common regulatory elements associated with these two genes. We are currently generating the monomer-specific antibodies that will allow us to address this question.

The topology of the trees in Figure 2, with Arabidopsis *CAT1* in a subclade with *Z. mays Cat2* and Arabidopsis *CAT3* in a subclade with *Z. mays Cat3*, further suggests that the duplication that gave rise to the progenitors of *CAT1* and *CAT3* predated the divergence of monocots and dicots from a common ancestor. Although this is an intriguing interpretation, it must remain speculative because not all catalases from each of the taxa have been analyzed, and it is equally plausible that this branching pattern represents an artifact caused by the limited gene sampling used in this analysis.

Baltimore (1985) postulated that reverse transcription of a processed cellular mRNA can generate intron-less cDNA copies of expressed genes that can be randomly integrated into the genome. The demonstration that the *S. cerevisiae* Ty elements transpose through an RNA intermediate (Boeke *et al.* 1985) established that retroelements normally present in the genome can provide a cellular source of reverse transcriptase; experimental demonstration that reverse transcription of a cellular mRNA provides a mechanism for intron loss
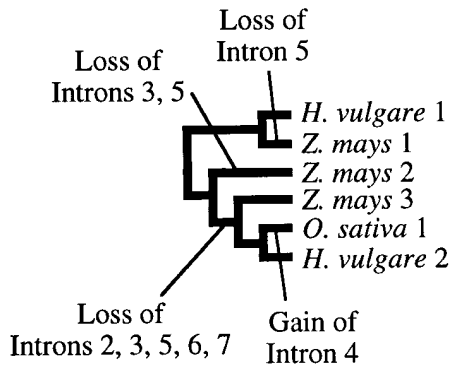
Figure 5.—A model for the evolution of intron structure within the grass subclade of catalases. We suggest that a mechanism of concerted intron loss in which reverse transcription of a cellular mRNA, followed by partial gene replacement of the endogenous gene copy by an intronless cDNA, is a relatively common event during genomic evolution and has occured at least twice in the evolution of the grass catalase genes. The gene duplications to create three genes, each containing seven introns, occurred early in the evolution of angiosperms and preceded the split between monocots and dicots. Subsequently, one event of concerted intron loss of introns 2–7 occurred in one grass subclade that includes *Z. mays Cat1*, *O. sativa CAT1*, and *H. vulgare CAT2*. After this event, the gain of an intron in position 4 occurred in the *O. sativa CAT1* sequence. An independent event of concerted intron loss of introns 3 and 5 occurred in *Z. mays Cat2*. The loss of a single intron in *Z. mays Cat1* could represent a third example of this mechanism of intron loss, but it could also have occurred via another mechanism.

when the cDNA copy replaces the endogenous genomic copy through homologous recombination has been provided in yeast (Derr *et al.* 1991). Introns located at the gene termini are less likely to be replaced by such a mechanism than are introns in the middle of the gene because little of the cDNA intermediate would extend beyond the intron, providing a less efficient substrate for homologous recombination. Moreover, this is a simple mechanism by which contiguous blocks of introns can be lost in one event. This argument has been invoked to explain the asymmetric location (at the gene termini, usually the 5′ end) of those few introns retained in the *S. cerevisiae* genome (Fink 1987) and in the red algae (Liaud *et al.* 1995).

One particularly intriguing example of multiple intron loss apparently occurred in one subclade of grass catalase genes (*Z. mays Cat3*, *O. sativa CAT1*, and *H. vulgare CAT2*). This subclade was defined on the basis of amino acid sequence similarity (Figure 2), and is well supported by the bootstrap analysis. Members of this subclade apparently have lost introns 2, 3, 5, 6, and 7, but they retain the 5′-most and 3′-most introns (Figure 5). This pattern of intron loss is consistent with reverse transcription of a cellular RNA, followed by gene replacement by a homologous recombination event in which the recombination break points lie downstream of the 5′-most intron and upstream of the 3′-most in-

tron. Such a mechanism of concerted loss of adjacent introns provides the most parsimonious explanation for the loss of multiple contiguous introns. Plants are rich in retroelements, which presumably could provide a cellular source of reverse transcriptase. For example, more than 60% of the DNA from a 280-kb region surrounding the maize *Adh1*-F locus represented retroelements (SanMiguel *et al.* 1996). Arabidopsis has more than 20 characterized retroelements, and there is evidence that at least some of these retroelements have been active since the founding of the Arabidopsis lineage and the divergence of the various ecotypes (Konieczny *et al.* 1991; Voytas and Ausubel 1988; Wright *et al.* 1996).

This mechanism of homologous recombination of a reverse-transcribed cDNA copy of a processed mRNA could also explain the loss of the adjacent introns 3 and 5 in the *Z. mays Cat2* sequence (Figure 3). Although the related *Z. mays Cat1* sequence also has lost intron 5, we infer that *Z. mays Cat1* and *Cat2* suffered independent losses of intron 5 (Figure 5). The most parsimonious explanation, based on the model of concerted intron loss, is that *Z. mays Cat2* lost introns 3 and 5 simultaneously and that *Z. mays Cat1* subsequently lost intron 5 (note that *H. vulgare CAT1* retains intron 5). Within the angiosperm catalases, there are no examples of loss of two nonadjacent introns without the loss of intervening introns. This mechanism of reverse transcription followed by homologous recombination could also explain the loss of single introns from Arabidopsis *CAT1* and *CAT3*, and has been invoked to explain the loss of single introns in potato and tomato actin genes (Drouin and Moniz de Sá 1997). Other examples of the loss of single introns from plant genes have been noted in a number of plant gene families (Huang *et al.* 1990; Kumar and Trick 1993; Häger *et al.* 1996). Other mechanisms, however, may also have been responsible for these losses of single introns. We suggest that the examples of concerted loss of multiple, contiguous introns in the grass catalases provide stronger evidence for gene replacement with a reverse-transcribed cDNA as a mechanism of intron loss.

Within the subclade of grass catalases in which introns 2, 3, 5, 6, and 7 have been lost, one sequence, *O. sativa CAT1*, contains an intron in a novel position (intron 4; Figures 3 and 5). The simplest explanation is that the *O. sativa CAT1* sequence gained an intron after the separation of the *O. sativa* lineage from the other grasses. The opposite explanation, the loss of intron 4 from all other plant catalases, would require multiple loss events in the individual lineages. Intron insertion has been postulated to have occurred in families of G protein genes (Dietmaier and Fabry 1994), in the triose-phosphate isomerase gene family (Kwiatkowski *et al.* 1995; Logsdon *et al.* 1995), and in the *RBCS* gene family in the Solanaceae (Fritz *et al.* 1993). On the basis of an analysis of the actin and tubulin gene families,

Dibb and Newman (1989) postulated that intron gain occurs between the G and R of the intron-flanking consensus sequence, C/AAGR, which they term a proto-splice site. This consensus corresponds to the consensus exon/intron 5′ splice junction and, therefore, is observed at the site of existing introns. Such sequences, however, also may originate and persist in the absence of introns because of coding constraints (Dibb and Newman 1989). In each of the grass catalases, the sequence surrounding the potential intron 4 site is CAGR, which corresponds exactly to the consensus site, yet only the *O. sativa CAT1* sequence has an intron at this position. We argue that this represents an example of intron gain at a proto-splice site.

We suggest that a mechanism of concerted intron loss in which reverse transcription of a cellular mRNA followed by partial gene replacement of the endogenous gene copy by an intronless cDNA is a relatively common event during genomic evolution. The evolution of the plant catalase gene family provides evidence for at least two independent events of intron loss by this mechanism. The gene duplications to create three genes occurred early in the evolution of angiosperms and preceded the split between monocots and dicots that occurred between 200 and 100 mya (Stewart and Rothwell 1993). Subsequently, one event of concerted intron loss of introns 2–7 occurred in one grass subclade that includes *Z. mays Cat1*, *O. sativa CAT1*, and *H. vulgare CAT2* (Figure 5). After this event, the gain of an intron in position 4 occurred in the *O. sativa CAT1* sequence. A second event of concerted intron loss occurred in *Z. mays Cat2*, and numerous other single-intron losses have occurred within the angiosperm catalases. In plants, examples of critical regulatory elements residing within intron sequences have accumulated in recent years (Callis *et al.* 1987; Fu *et al.* 1995a,b; Kao *et al.* 1996; Sieburth and Meyerowitz 1997). Intron loss or gain, as well as the modification of sequences residing within introns, may therefore provide critical steps in the divergence of expression patterns of individual gene family members.

LITERATURE CITED

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman *et al.*, 1997 *Current Protocols in Molecular Biology.* Greene Publishing Associates/Wiley Interscience, NY.

Baltimore, D., 1985 Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. Cell **40:** 481–482.

Beevers, H., 1982 Glyoxysomes in higher plants. Ann. NY Acad. Sci. **386:** 243–253.

Beyer, W. F., Jr., and I. Fridovich, 1987 Catalases-with and without heme, pp. 651–661 in *Oxygen Radicals in Biology and Medicine*, edited by M. G. Simic, K. A. Taylor, J. F. Ward and C. Von Sonntag. Plenum Press, NY.

Boeke, D., D. J. Garfinkel, C. A. Styles and G. R. Fink, 1985 Ty elements transpose through an RNA intermediate. Cell **40:** 491–500.

Callis, J., M. Fromm and V. Walbot, 1987 Introns increase gene expression in cultured maize cells. Genes Dev. **1:** 1183–1200.

Canvin, D. T., 1990 Photorespiration and CO_2 concentrating mechanisms, pp. 253–273 in *Plant Physiology, Biochemistry and Molecular Biology*, edited by D. T. Dennis and D. H. Turpin. Longman Scientific and Technical, Harlow, UK.

Choi, S. D., R. Creelman, J. Mullet and R. A. Wing, 1995 Construction and characterization of a bacterial artificial chromosome library from *Arabidopsis thaliana.* Weeds World **2:** 17–20.

Clegg, M. T., M. P. Cummings and M. L. Durbin, 1997 The evolution of plant nuclear genes. Proc. Natl. Acad. Sci. USA **94:** 7791–7798.

Cohen, G., W. Rapatz and H. Ruis, 1988 Sequence of the *Saccharomyces cerevisiae CAT1* gene and amino acid sequence of catalase A derived from it. Eur. J. Biochem. **176:** 159–163.

Dietmaier, W., and S. Fabry, 1994 Analysis of the introns in genes encoding small G proteins. Curr. Genet. **26:** 497–505.

Derr, L. K., J. N. Strathern and D. J. Garfinkel, 1991 RNA-mediated recombination in *S. cerevisiae.* Cell **67:** 355–364.

Dibb, N. J., and A. J. Newman, 1989 Evidence that introns arose at proto-splice sites. EMBO J. **8:** 2015–2021.

Didion, T., and R. Roggenkamp, 1992 Targeting signal of the peroxisomal catalase in the methylotrophic yeast *Hansenula polymorpha.* FEBS Lett. **303:** 113–116.

Drouin, G., and M. Moniz de Sá, 1997 Loss of introns in the pollen-specific actin gene subfamily members of potato and tomato. J. Mol. Evol. **45:** 509–513.

Feinberg, A. P., and B. Vogelstein, 1984 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137:** 266–267.

Fink, G. R., 1987 Pseudogenes in yeast? Cell **49:** 5–6.

Fowler, T., M. W. Rey, P. Vaha-Vahe, S. D. Power and R. M. Berka, 1993 The catR gene encoding a catalase from *Aspergillus niger*: primary structure and elevated expression through increased gene copy number and use of a strong promoter. Mol. Microbiol. **9:** 989–998.

Frugoli, J. A., H. H. Zhong, M. L. Nuccio, P. McCourt, M. A. McPeek *et al.*, 1996 Catalase is encoded by a multi-gene family in *Arabidopsis thaliana* (L.) Heynh. Plant Physiol. **112:** 327–336.

Fritz, C. C., F. P. Wolter, V. Schenkemeyer, T. Herget and P. H. Schreier, 1993 The gene family encoding the ribulose-(1,5)-bisphosphate carboxylase/oxygenase (Rubisco) small subunit of potato. Gene **137:** 271–274.

Fryxell, K. J., 1996 The coevolution of gene family trees. Trends Genet. **12:** 356–369.

Fu, H., S. Y. Kim and W. D. Park, 1995a High-level tuber expression and sucrose inducibility of a potato *Sus4* sucrose synthase gene require 5′ and 3′ flanking sequences and the leader intron. Plant Cell **7:** 1387–1394.

Fu, H., S. Y. Kim and W. D. Park, 1995b A potato *Sus3* sucrose synthase gene contains a context-dependent 3′ element and a leader intron with both positive and negative tissue-specific effects. Plant Cell **7:** 1395–1403.

Guan, L., and J. G. Scandalios, 1996 Molecular evolution of maize catalases and their relationship to other eukaryotic and prokaryotic catalases. J. Mol. Evol. **42:** 570–579.

Häger, K.-P., B. Müller, C. Wind, S. Erbach and H. Fischer, 1996 Evolution of legumin genes: loss of an ancestral intron at the beginning of angiosperm diversification. FEBS Lett. **387:** 94–98.

Huang, N., T. D. Sutliff, J. C. Litts and R. L. Rodriguez, 1990 Classification and characterization of the rice α-amylase multigene family. Plant Mol. Biol. **14:** 655–668.

Kao, C., S. Cocciolone, I. Vasil and D. McCarty, 1996 Localization and interaction of the *cis*-acting elements for abscisic acid,

VIVIPAROUS1, and light activation of the *C1* gene of maize. Plant Cell **8:** 1171–1179.

Kendall, A. C., A. J. Keys, A. J. Turner, P. J. Lea and B. J. Merlin, 1983 The isolation and characterization of a catalase deficient mutant of barley (*Hordeum vulgare* L.). Planta **159:** 505–511.

Konieczny, A., D. F. Voytas, M. P. Cummings and F. M. Ausubel, 1991 A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics **127:** 801–809.

Kumar, V., and M. Trick, 1993 Sequence complexity of the *S* receptor kinase gene family in *Brassica.* Mol. Gen. Genet. **241:** 440–446.

Kwiatkowski, J., M. Krawczyk, M. Kornacki, K. Bailey and F. J. Ayala, 1995 Evidence against the exon thoery of genes derived from the triose-phosphate isomerase gene. Proc. Natl. Acad. Sci. USA **92:** 8503–8506.

Liaud, M.-F., U. Brandt and R. Cerff, 1995 The marine red alga *Chondrus crispus* has a highly divergent β-tubulin gene with a characteristic 5′ intron: functional and evolutionary implications. Plant Mol. Biol. **28:** 313–325.

Logsdon, J. M. J., M. G. Tyshenko, C. Dixon, D.-J. Jonathon, V. K. Walker *et al.,* 1995 Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. Proc. Natl. Acad. Sci. USA **92:** 8507–8511.

Low, P. S., and J. R. Merida, 1996 The oxidative burst in plant defense: function and signal transduction. Physiol. Plant. **96:** 533–542.

McClung, C. R., 1997 The regulation of catalase in *Arabidopsis.* Free Radical Biology & Medicine **23:** 489–496.

McDowell, J. M., S. Huang, E. C. McKinney, Y. Q. An and R. B. Meagher, 1996 Structure and evolution of the actin gene family in *Arabidopsis thaliana.* Genetics **142:** 587–602.

Meagher, R. B., 1991 Divergence and differential expression of actin gene families in higher plants. Int. Rev. Cytol. **125:** 139–163.

Mehdy, M. C., Y. K. Sharma, K. Sathasivan and N. W. Bays, 1996 The role of activated oxygen species in plant disease resistance. Physiol. Plant. **98:** 365–374.

Meyerowitz, E. M., and C. R. Somerville, 1994 *Arabidopsis.* Cold Spring Harbor Monograph Series. 1300.

Murray, W. W., and R. A. Rachubinski, 1987 The nucleotide sequence of complementary DNA and the deduced amino acid sequence of peroxisomal catalase of the yeast *Candida tropicalis* pK233. Gene **61:** 401–413.

Murray, W. W., and R. A. Rachubinski, 1989 Nucleotide sequence of peroxisomal catalase from the yeast *Candida tropicalis* pK233: identification of an upstream *Bam*HI polymorphism. Nucleic Acids Res. **17:** 3600.

Nakagawa, C. W., N. Mutoh and Y. Hayashi, 1995 Transcriptional regulation of catalase gene in the fission *Schizosaccharomyces pombe*: molecular cloning of the catalase gene and Northern blot analyses of the transcript. J. Biochem. **118:** 109–116.

Nakashima, H., M. Yamamoto, K. Goto, T. Osumi, T. Hashimoto *et al.,* 1989 Isolation and characterization of the rat catalase-encoding gene. Gene **79:** 279–288.

Navarro, R. E., M. A. Stringer, W. Hansberg, W. E. Timberlake and J. Aquirre, 1996 catA, a new *Aspergillus nidulans* gene encoding a developmentally regulated catalase. Curr. Genet. **29:** 352–359.

Ohta, T., 1991 Multigene families and their implications for evolutionary theory, pp. 15–25 in *New Aspects of the Genetics of Molecular Evolution*, edited by M. Kimuar and N. Takahata. Japan Scientific Societies Press: Tokyo.

Okada, H., M. Ueda, T. Sugaya, H. Atomi, S. Mozaffar *et al.,* 1987 Catalase gene of the yeast *Candida tropicalis.* Sequence analysis and comparison with peroxisomal and cytosolic catalases from other sources. Eur. J. Biochem. **170:** 105–110.

Orr, E. C., G. C. Bewley and W. C. Orr, 1990 cDNA and deduced amino acid sequence of *Drosophila* catalase. Nucleic Acids Res. **18:** 3663.

Orr, W. C., E. C. Orr, S. K. Legan and R. S. Sohal, 1996 Molecular analysis of the *Drosophila* catalase gene. Arch. Biochem. Biophys. **330:** 251–258.

Quan, F., R. G. Korneluk, M. B. Tropak and R. A. Gravel, 1986 Isolation and characterization of the human catalase gene. Nucleic Acids Res. **14:** 5321–5335.

Ryals, J. A., U. H. Neuenschwander, M. G. Willits, A. Molina, H. Y. Steiner *et al.,* 1996 Systemic acquired resistance. Plant Cell **8:** 1809–1819.

SanMiguel, P., A. Tikonov, Y. K. Jin, N. Motchoulskaia, D. Zakharov *et al.,* 1996 Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765–768.

Scandalios, J. G., 1987 The antioxidant enzyme genes *Cat* and *Sod* of maize: regulation, functional significance, and molecular biology, pp. 19–43 in *Isozymes: Current Topics in Biological and Medical Research*, edited by M. C. Rattazi, J. G. Scandalios and G. S. Whitt. Alan R. Liss, Inc, NY.

Shaffer, J. B., K. E. Preston and B. A. Shepard, 1990 Nucleotide and deduced amino acid sequences of mouse catalase: molecular analysis of a low affinity mutant. Nucleic Acids Res. **18:** 4941.

Sieburth, L. E., and E. M. Meyerowitz, 1997 Molecular dissection of the *AGAMOUS* control region shows that *cis* elements for spatial regulation are located intragenically. Plant Cell **9:** 355–365.

Somerville, C. R., and W. L. Ogren, 1982 Genetic modification of photorespiration. Trends Biochem. Sci. **17:** 171–174.

Spevak, W., A. Hartig, P. Meindl and M. Ruis, 1986 Heme control region of the catalase T gene of the yeast *Saccharomyces cerevisiae.* Mol. Gen. Genet. **203:** 73–78.

Stewart, W. N., and G. W. Rothwell, 1993 *Paleobotany and the Evolution of Plants.* Cambridge University Press: Cambridge, UK.

Subramani, S., 1993 Protein import into peroxisomes and biogenesis of the organelle. Annu. Rev. Cell Biol. **9:** 445–478.

Suzuki, M., T. Ario, T. Hattori, K. Nakamura and T. Asahi, 1994 Isolation and characterization of two tightly linked catalase genes from castor bean that are differentially regulated. Plant Mol. Biol. **25:** 507–516.

Swofford, D. L., 1993 *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1.* Illinois Natural History Survey, Champaign, IL.

Trelease, R. N., 1984 Biogenesis of glyoxysomes. Annu. Rev. Plant Physiol. **35:** 321–347.

von Ossowski, I., G. Hausner and P. C. Loewen, 1993 Molecular evolutionary analysis based on the amino acid sequence of catalase. J. Mol. Evol. **37:** 71–76.

Voytas, D. F., and F. M. Ausubel, 1988 A copia-like transposable element family in *Arabidopsis thaliana.* Nature **336:** 242–244.

Waterston, R., C. Martin, M. Craxton, C. Huynh, A. Coulson *et al.,* 1992 A survey of expressed genes in *Caenorhabditis elegans.* Nature Genet. **1:** 114–123.

Willekens, H., C. Langebartels, C. Tire, M. van Montagu, D. Inzé *et al.,* 1994a Differential expression of catalase genes in *Nicotiana plumbaginifolia* (L.). Proc. Natl. Acad. Sci. USA **91:** 10450–10454.

Willekens, H., R. Villarroel, M. van Montagu, D. Inzé and W. van Camp, 1994b Molecular identification of catalases from *Nicotiana plumbagnifolia* (L.). FEBS Lett. **352:** 79–83.

Wilson, R., R. Ainscough, K. Anderson, C. Baynes, M. Berks *et al.,* 1994 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans.* Nature **368:** 32–38.

Wright, D. A., N. Ke, J. Smalle, B. M. Hauge, H. M. Goodman *et al.,* 1996 Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana.* Genetics **142:** 569–578.

Yang, Y., J. Shah and D. F. Klessig, 1997 Signal perception and transduction in plant defense responses. Genes Dev. **11:** 1621–1639.

Yaun, H. T., C. D. Bingle and F. J. Kelly, 1996 Differential patterns of antioxidant enzyme mRNA expression in guinea pig lung and liver during development. Biochim. Biophys. Acta **1305:** 163–171.

Zhong, H. H., and C. R. McClung, 1996 The circadian clock gates expression of two *Arabidopsis* catalase genes to distinct and opposite circadian phases. Mol. Gen. Genet. **251:** 196–203.

Zhong, H. H., J. C. Young, E. A. Pease, R. P. Hangarter and C. R. McClung, 1994 Interactions between light and the circadian clock in the regulation of *CAT2* expression in *Arabidopsis.* Plant Physiol. **104:** 889–898.

Zhong, H. H., A. S. Resnick, M. Straume and C. R. McClung, 1997 Effects of synergistic signaling by phytochrome A and cryptochrome 1 on circadian clock-regulated catalase expression. Plant Cell **9:** 947–955.